We thank the reviewers and editors for their feedback. In addition to their comments, we made the following changes:

- Corrected Fig. S11d, which had the wrong dataset plotted for segment 2–75 Gy.
- Corrected the files in the associated supplemental data for sample Gi326.

Associate editor

Section 3.1 and 3.2: Initial signal rise
This gets me back to the old question we were asking, why it never not happened in the measurements of Krbetschek's lab. Do you think it is due to their preperation technique (flotation)?

Reply: We did a preliminary test on a sample that had undergone flotation (not yet published) and also observed the initial rise, so it is probably not an issue of sample preparation only. Currently, our hypothesis is that the difference was caused by the detection systems (filter+PMT vs. spectrometer), though we do not have the data to prove this. It is our understanding that the expected transmission of filters is based on a normal incidence (i.e., 90° to the filter surface) and that other angles shift the transmission to shorter wavelengths (e.g., https://www.edmundoptics.com/knowledge-center/application-notes/optics/optical-filter-orientation/?srsltid=AfmBOorEUmySJtiMhXOV3G9vc6xriml0lJ3Y1MZ9li48DBzTxPlFnaO8). So, there may be some stray transmission of highly angled emissions outside the desired detection window. Using a spectroscopic detection, this issue would be suppressed. Below, we have added a very rough comparison of both detection systems for the same samples from previously published data. It is by no means definitive, but the initial rise seems to be somewhat reduced in the spectrometer system (note that the channel lengths were twice as long for the spectrometer measurement).

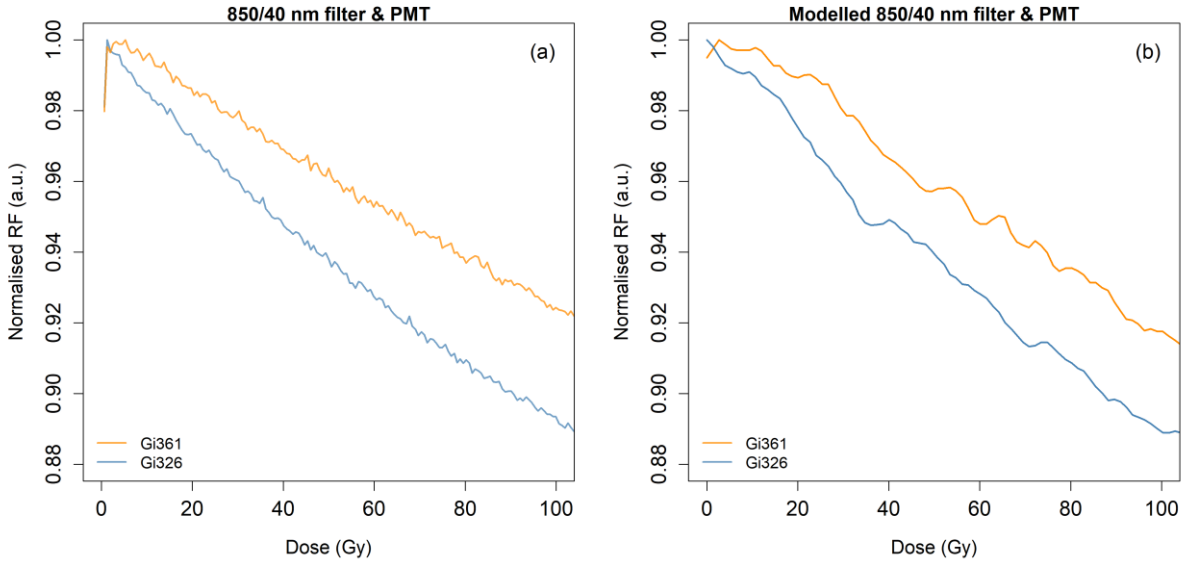

Fig. 1: Regenerative dose response curves obtained using (a) a PMT and an 850/40 nm filter or (b) a spectrometer, modelling the transmitted signal according to the filter and PMT transmission datasheets. RF is normalised to the highest signal intensity. Different aliquots were used for each panel. The underlying data are shown in Sontag-González and Fuchs (2022) (a) Fig. 1c and (b) Fig. S1 g–h.

Section 5 MAR
Even after following the discussion between the authors and reviewers, the MAR results seem to be

most promissing. And you clearly understand the logic behind it. Personally I would use the MAR data to be the representative of IR-RF in Fig. 6b and Fig. S13. (but feel free to disagree!)

Reply: Thanks! It's very encouraging to hear that. We have added the MAR data to Fig. S13 for completion but take a more conservative approach to Fig. 6b, as we feel the method is still too preliminary to be considered the most reliable of the IR-RF methods. In particular, we still see a change in D_e with progressively longer segments, indicating that the issue is not solved. For better comparability, we have also added a new figure to the supplement (Fig. S15) with a summary of the change in MAR D_e with segments equivalent to Fig. 10 for the SAR results.

Tables 3 and 4
Please use a consistent significant number of digits. Especially the inconsistency between the IR-RF and IRPL data do not look good (up to 4 digits in IR-RF and up to 3 digits in IRPL except Gi326).

Reply: We reduced the number of significant digits to match the IRPL data (taken directly from Kumar et al. (2021)).

Anonymous referee #1

 # General remarks

The authors have substantially revised their manuscript and made significant efforts to address all suggestions. The findings are supported by the data and the manuscript follows established standards for data transparency and reporting.

Therefore, I recommend the publication of the manuscript in Geochronology, pending the incorporation of a few comments I have listed below (naturally do authors do not have to agree to all of them). Although I have now flagged some minor issues I had overlooked the last time, I do not need to see the manuscript again.

 # General comments

* I had overlooked that last time that the authors have reported expected ages in Table 1, but never reported the comparison (IR-RF/IRPL vs expected) ages. Please add a table or plot for this; as for the approach, the authors can pick the one they prefer; however, it should be consistent. If it does not fit the main text, please report it in the supplement and refer to it. I had stumbled over it because the authors mention in their abstract and age span but then only work with equivalent doses, while the readers would be most likely more interested in the final age comparison.
Reply: Our findings are expected to relate only to the equivalent dose (i.e., irrespective of sample dose rate), so we have added the expected doses to Table 1 and rephrased the sentence in the abstract to highlight the sample doses rather than the ages: "For four out of the seven tested known-age samples spanning ca. 100–300 Gy (20–130 ka), we obtained results in keeping with the expected doses. Two additional modern samples, however, yielded slight dose underestimations."

 # Response to author's responses

> We have commented on this issue in section 2.1, but since all samples expected to be
> in the dating range received the same treatment, we do not expect etching to affect our
> conclusions.

Perhaps you can rephrase the 2nd part of your addition in Sec. 2.1? I find it confusing that you connect the 'dating range' with the 'same treatment' to conclude that your results are not affected. I think the reader can guess what you mean, but it is certainly not straightforward to understand.
Reply: We have rephrased for clarity: "The effect on the resulting D_e is poorly studied, but all samples expected to be in the dating range (i.e., not saturated) received the same treatment (HF etching), so any variation in D_e accuracy we observe for these samples would not be caused by a difference in sample preparation. Thus, we do not expect etching to affect our conclusions."

> In any case, we have rephrased this sentence to "a more athermally stable signal".

The information provided deviates from the content of Krbetschek et al. (2000) and should not be attributed to the authors. Please adhere to the original text or provide a different reference.
Reply: We have rephrased to: "The main advantages of IR-RF dating over the more common infrared stimulated luminescence (IRSL) of K-feldspar (Hütt et al., 1988) include a more athermally stable signal (based on IR-RF fading tests by Krbetschek et al., (2000) suggesting signal stability)…" according to the statement by Krbetshek et al. (2000): "Fading tests (storage over periods of several months at room temperature) have shown signal stability" (p. 497).

> We did not state the number of rejected channels here in the methods section because
> this was a parameter we varied. We have added a sentence stating that between 0 and
> 499 channels were ignored, which will be [...]

Those figures and experiments you have produced are excellent, and I suggest that you add them to your supplementary material (except for the one you have in the main text anyway) because readers will unlikely look up all the discussion.
Reply: Thank you! We have added the dose vs. signal figure to the supplement.

Detailed comments

My line numbers refer to the version with the changes tracked.

* L24: I think that, given the current understanding of IR-RF, it is also trap specific and the authors seem to confirm this multiple times in the manuscript. Please rephrase.
Reply: We have rephrased and added a separate sentence: "Like IR-RF, IRPL is also expected to be trap-specific."

* L22-L24: Given the IRPL appears in the title of your manuscript and is an essential part, the phrase feels oddly formulated. Please prioritise it according to your manuscript content.
Reply: Our results in the manuscript are more heavily focussed on IR-RF than IRPL dating, so we feel that the abstract is correctly proportioned between the two methods. Unfortunately, this proportion is more difficult to portray in the title.

* L39: Add proper reference for the SAR approach and more correctly you should refer to the IRSAR approach (Erfurt and Krbetschek, 2003) where applicable as this approach does not use a test dose for sensitivity correction, while it is an essential parameter in the original SAR approach by Murray

and Wintle (2000)
Reply: True, that is an important distinction. We have added the reference: “(IRSAR; Erfurt and Krbetschek, 2003)

* L89: I had somewhat overlooked this the last time. What makes you believe that 'signal instability' is the primary reason for the observed saturation? Out of the many possibilities, this is the one with the lowest explanatory power, given that we can indeed successfully date events using IR-RF and IRPL. If the principle trap is indeed unstable, this shouldn't be possible.

Reply: We have removed the suggestion.

* L125: Subtle issue: Autzen et al. (2022) reported the corrections, Tribolo et al. (2019) flagged the 'issue'.
Reply: We have changed the reference.

* L129: Minor wording inconsistency: The factor you applied reads 1.0825.
Reply: Thanks for catching that! Corrected.

* L135: Also here, 8.25\% can mean reduction or a boost.
Reply: Corrected, as above.

* L211: Why over four million? The formula should be n(n-1)/2 -> 1500*1499/2 -> 1,124,250 (ordered permutations for a segment length of 2 channels). Besides, I suggest removing the sentences starting from 'Since [...]' they read verbose and add little to manuscript.

Reply: We have removed the sentence, as suggested, but we note that we had 3000 channels (so, 3000*2999/2=4498500).

* L226: I suggest removing 'These results [...]' because it is indeed a suggestive statement that is certainly true for all multi-grain luminescence measurements but it depends on many factors unrelated to the 'multi-grain' nature.
Reply: It is true that variation at the single-grain level can lead to issues at the multi-grain level for all multi-grain luminescence measurements, so we have changed the sentence to specify that we mean only IR-RF measurements. However, we have chosen to retain the sentence, as it introduces the modelling in the rest of the paragraph.

* L230: I propose removing this sentence, as it could potentially be misinterpreted and used as a justification odd data treatments. By removing it, you can avoid the risk of it being misinterpreted as a cure for the symptoms rather than using remedy, for instance, better sample preparation and potentially signal deconvolution. If you wish to retain the phrase, it is essential to clearly state the problems associated with it. Specifically, any removal of initial parts of the curve is somewhat arbitrary and should not be used to “tune” the results.
Reply: We have retained the observation because it is the basis for some investigations in later segments but we have rephrased the conclusions to avoid misinterpretation: “This suggests that (i) the ‘initial signal rise’ originates from signal contamination by presumably non-K-feldspar minerals and (ii) the DRC of the modelled ‘contaminated’ aliquot converges with that of the ‘pure’ one (for this sample at ~100 Gy).”

* L595 onwards: Please repeat in one sentence the mission of your work before jumping the summary. Furthermore, to help the reader grasping all your results, you should bullet point your investigations and then conclude the outcome in one, maximum two sentences per point.
Reply: We have added a short summary of our intent before the results summary, as suggested: "We tested whether the methodological developments of the past decade have improved the accuracy of IR-RF dating of known age samples which had previously yielded inaccurate IR-RF ages with an IRSAR protocol. Specifically, we re-analysed previous data and re-measured samples using improved measurement and data analysis protocols (i.e., increased measurement temperature and vertical sliding) as well as using new methods (i.e., MAR IR-RF and IRPL)." However, we chose to retain the full-text version of the results summary.

* L659-L661: Journal and DOI missing.
Reply: Added.

* L682-L684: DOI missing.
Reply: Added.

* L691-L693: DOI missing.
Reply: Added URL.

* L732: Canonical URL entry missing
Reply: Added.

Tables and figures

* Table 1: Where the age is around zero (055642, 102011) you should write ~0 or <0.1 ka; the quoted negative age (in particular with that precision) does not make sense.
Reply: We have changed it to "ca. 0".

Comments supplement

* Figure S2. This plot illustrates that a KDE may not always be meaningful if individual uncertainties are significant (in particular: Gi326). It also demonstrates that the quoted uncertainties in Table 4 are misleading for the given sample. The individual uncertainties are important (except of sample H22553) and should not be disregarded. Regardless of previous studies, it is crucial to provide meaningful uncertainties supported by the data.
Reply: We thank the reviewer for this comment. Indeed, the uncertainty of Gi326 in Fig. S2 did not match that reported in Table 4. We double-checked the data and found we had forgotten to include the 5% source calibration to the uncertainty of this sample (for the other samples it had been included). We have now corrected Table 4 and Figs. 5,7,9 and S9 to include the correct uncertainties. Additionally, we agree that uncertainty-weighted central estimates are an important metric. However, since the focus of this study was the comparison with published values rather than dating, and considering that we did not have the underlying data from the previously published IRPL results, we think the reported uncertainties (after the correction) are sufficient.

* Figure S13: In particular the Abanico plot is rather essential and should not be hidden in the supplement but appear in the main text. If you don't want to remove a main figure, add it to Fig. 6.

Reply: We have added the abanico plot as panel c to Fig. 6, as suggested. Note that this version is slightly different because we noticed we had not used the corrected expected D_e values (source dose rate calibration issue) in the first version.

References

Autzen, M., Andersen, C.E., Bailey, M., Murray, A.S., 2022. Calibration quartz: An update on dose calculations for luminescence dating. Radiation Measurements 106828. https://doi.org/10.1016/j.radmeas.2022.106828

Erfurt, G., Krbetschek, M.R., 2003. IRSAR - A single-aliquot regenerative-dose dating protocol applied to the infrared radiofluorescence (IR-RF) of coarse-grain K-feldspar. Ancient TL 21, 35–42. https://doi.org/10.26034/la.atl.2003.358

Murray, A.S., Wintle, A.G., 2000. Luminescence dating of quartz using an improved single-aliquot regenerative-dose protocol. Radiation Measurements 32, 57–73. https://doi.org/10.1016/s1350-4487(99)00253-x

Tribolo, C., Kreutzer, S., Mercier, N., 2019. How reliable are our beta-source calibrations? Ancient TL 37, 1–10. https://doi.org/10.26034/la.atl.2019.529

###

Sontag-González, M. and Fuchs, M.: Spectroscopic investigations of infrared-radiofluorescence (IR-RF) for equivalent dose estimation, Radiation Measurements, 153, 106733, https://doi.org/10.1016/j.radmeas.2022.106733, 2022.